Am. J. Hum. Genet. 41:1104–1116, 1987

Maximum-Likelihood Estimation of Familial Correlations from Multivariate Quantitative Data on Pedigrees: A General Method and Examples

D. C. RAO,*,† GEORGE P. VOGLER,* M. MCGUE,*,‡ AND J. M. RUSSELL*

*Division of Biostatistics and †Departments of Psychiatry and Genetics, Washington University School of Medicine, St. Louis; and ‡Department of Psychology, University of Minnesota, Minneapolis

SUMMARY

A general method for maximum-likelihood estimation of familial correlations from pedigree data is presented. The method is applicable to any type of data structure, including pedigrees in which variable numbers of individuals are present within classes of relatives, data in which multiple phenotypic measures are obtained on each individual, and multiple group analyses in which some correlations are equated across groups. The method is applied to data on high-density lipoprotein cholesterol and total cholesterol levels obtained from participants in the Swedish Twin Family Study. Results indicate that there is strong familial resemblance for both traits but little cross-trait resemblance.

INTRODUCTION

Correlations among relatives are often used to measure the degree of familial aggregation of a certain trait such as cholesterol. Information regarding the genetic and environmental sources of familial resemblance is provided by such correlations. Familial correlations can be of several types: (1) correlations between classes of relatives on the same variable, such as correlations between cholesterol levels of two spouses or of a parent and child, (2) correlations between multiple members of the same class of relatives on the same variable, such as the sibling cholesterol correlation, (3) correlations between classes of

Received July 24, 1986.

Address for correspondence and reprints: Dr. D. C. Rao, Division of Biostatistics, Washington University School of Medicine, 660 South Euclid, Box 8067, St. Louis, MO 63110.

relatives on different variables measured in each class, such as the correlation between cholesterol in a parent and high-density lipoprotein cholesterol (HDL) in a child, (4) correlations between members of the same class on different variables measured on each member, such as the correlation between cholesterol in a child and HDL in a sibling, and (5) correlations between two different variables measured on the same individual.

In most applications, pedigrees are not of uniform size. In 2-generational nuclear families, the sibship size can vary, and in extended pedigrees there is even greater variation among family configurations. Consequently, conventional ways of computing interclass and intraclass correlations are not readily applicable with family data.

Estimation of correlations in complex pedigrees has thus far been addressed only for specific situations. Maximum-likelihood methods have been developed for (1) the estimation of interclass correlations when a single observation is obtained on one class, such as a single parent, and multiple observations are obtained on a second class, such as children (Rosner 1979); (2) simultaneous estimation of 16 correlations from nuclear families in which a measure of the phenotype and an index of the environment are obtained on each individual (Rao et al. 1979); (3) the general case of interclass correlation when a variable number of observations occur in each class, such as one or both parents and varying sibship size in the offspring (Rao et al. 1979; Rosner 1982); (4) intraclass correlation within a class of varying size (Donner and Koval 1980); and (5) a sequential method for estimating familial interclass and intraclass correlations (Rao et al. 1982).

In this report, a general method is presented for obtaining maximum-likelihood estimates of familial correlations from quantitative data on pedigrees. The method can be used to estimate familial correlations either directly or as functions of the parameters of a hypothesized model of familial resemblance—for example, a model of genetic and cultural transmission with assortative mating in nuclear families. Furthermore, the method is sufficiently general to be applied to the analysis of any family configuration for which data are obtained on any number of phenotypic measures for each individual. Thus, all previous approaches to the maximum-likelihood estimation of familial correlations are special cases of this general method.

METHODS

For simplicity, the method is developed first for the univariate case, in which one phenotypic measure is obtained on each individual. It is assumed that the data are suitably adjusted for the effects of concomitant variables such as age and sex prior to applying this methodology. Consider the ith of N randomly sampled pedigrees. Let the column vector of adjusted phenotypes, one on each individual, be denoted by the vector x_i, where x_{ijk} denotes the phenotypic measure obtained on the kth individual in the jth group of the ith pedigree. The jth group ($j = 1, 2, \ldots, n_i$) is a class of relatives within a pedigree having an intraclass structure—for example, in a nuclear family $j = 1$ may indicate the father, $j = 2$ the mother, and $j = 3$ all offspring. The number of individuals

within a group ($k = 1, 2, \ldots, s_{ij}$) can vary among groups as well as among family units; in the above example, s_{i1} for fathers and s_{i2} for mothers can equal 0 or 1, and s_{i3} indicates the number of offspring in family i. Assume that x_i follows a multivariate normal distribution $N_{p_i}(\mu_i, \Sigma_i)$ with the dimensionality of x_i defined as

$$p_i = \sum_{j=1}^{n_i} s_{ij} \ .$$

The column vector μ_i contains subvectors of means for each group ($\mu_{i1}, \mu_{i2}, \ldots, \mu_{in_i}$), where each subvector is of dimension s_{ij} ($j = 1, 2, \ldots, n_i$) and the means for all individuals within a group are all assumed to be equal—for example, all elements of μ_{i1} are equal. It is further assumed that, within a group, all individual observations have the same variance and all pairwise observations have the same correlation (i.e., intraclass structure). Likewise, it is assumed that correlations between pairs of observations, one from each of two groups, are equal for all pairs from the two groups (i.e., interclass structure). Thus, the correlational structure is intraclass within groups and interclass between groups. The variance-covariance matrix Σ_i is defined as $\sigma_i \mathbf{R}_i \sigma_i$, where σ_i is a $(p_i \times p_i)$ diagonal matrix of SDs in which the first s_{i1} diagonal elements consist of the SD for group 1, the next s_{i2} elements are the SD for group 2, and so on. The matrix $\mathbf{R}_i$ is a correlation matrix consisting of n_i^2 submatrices of dimension ($s_{ij} \times s_{ik}$; $j = 1, 2, \ldots, n_i$; $k = 1, 2, \ldots, n_i$), where the diagonal submatrices contain intraclass correlations within groups and the off-diagonal submatrices consist of interclass correlations between groups.

The log likelihood for the ith family is given by

$$\ln L_i = -\tfrac{1}{2}[\ln|\Sigma_i| + (x_i - \mu_i)'\Sigma_i^{-1}(x_i - \mu_i)] + \text{constant} \ ,$$

where $|\Sigma_i|$ is the determinant, Σ_i^{-1} is the inverse of Σ_i, and $(x_i - \mu_i)$ is a column vector of deviations of the observed data from the group means for each observation in the ith family. The overall log likelihood for N families is

$$\ln L = \sum_{i=1}^{N} \ln L_i \ .$$

Although the correlations in the overall **R** matrix can be estimated directly using this method, it is sometimes useful to express the correlations as functions of parameters under a hypothesized model for the sources of familial resemblance, such as a model of genetic and cultural transmission and assortative mating (see, e.g., Rao et al. 1982). Maximum-likelihood estimates of the parameters (the vector θ of model parameters or the correlations **R** in addition to the means and variances) can then be obtained by numerically maximizing the log likelihood (e.g., Kaplan and Elston 1972; Lalouel 1979) with respect to the parameters.

When correlations are estimated directly, the maximum number of unknown parameters is n means, n variances, n intraclass correlations, and $n(n - 1)/2$ interclass correlations, for a maximum total of $n(n + 5)/2$, where n indicates the maximum number of groups in any pedigree. In most applications, there are fewer parameters to estimate than the maximum. For example, in nuclear families there are three groups of relatives (fathers, mothers, and offspring), so the maximum number of parameters is 12. However, there is no intraclass correlation in the groups consisting of mothers or fathers, so there are only 10 possible parameters. Furthermore, without noticeable change in precision, the means and variances can be fixed at their sample (i.e., moment) estimates for the two parental groups, further reducing the number of parameters to six. In certain situations, biological considerations would require that some of the intraclass and/or interclass correlations be equal, reducing the number of estimated parameters even further. For example, although each sibship in an extended pedigree will constitute a distinct group, one would ordinarily require that the sibling correlation be the same in every sibship. Similar considerations will apply to interclass correlations such as parent-child and uncle-niece.

The maximum-likelihood estimation procedure can be extended to the analysis of multiple phenotypic measures obtained on each individual. This is accomplished by expanding the univariate formulation to an additional dimension, so that with q variables per individual, each element of the **R** matrix in the univariate case becomes a $(q \times q)$ submatrix of **R** in the multivariate case, each element of μ becomes a $(q \times 1)$ vector of means, and each diagonal element of σ is reformulated as a diagonal submatrix containing q SDs (one for each phenotype within the group).

When multiple phenotypic measures are analyzed, in addition to estimates of interclass and intraclass correlations for each phenotype, the following types of cross-correlations are estimated: correlations between groups of relatives on different phenotypes, correlations between members within a group on different phenotypes, and correlations among different phenotypes within individuals. Therefore, there are generalized intraclass covariance structures within groups and generalized interclass covariance structures among groups, of which the univariate design is a special case. When all measures are standardized, the generalized intraclass covariance structure within a group consists of two components: (1) interclass correlations among measures within individuals and (2) correlations that exist between pairs of individuals and are intraclass for the same phenotype and interclass for different phenotypes; the first component is the multivariate generalization of the standardized variance of unity in the univariate case, and the second component is the multivariate generalization of the simple intraclass correlation. Similarly, the generalized interclass covariance structure consists of interclass correlations between individuals in different groups on all pairwise combinations of phenotypes.

The maximum number of parameters in the general multivariate case is qn means (where q is the number of variables and n is the maximum number of classes), qn variances, and $q^2n(n + 1)/2$ correlations of the following types: $nq(q + 1)/2$ intraclass correlations, $q^2n(n - 1)/2$ interclass correlations, and

$nq(q - 1)/2$ self-correlations within individuals and within groups. In the bivariate case for nuclear families with three groups, there is a maximum of six means, six variances, and 24 correlations. By using sample means and variances for both variables of mothers and fathers, and because there are no intraclass correlations in the groups consisting of mothers only or fathers only, the number of estimated parameters is reduced to 22.

This method of *m*aximum *l*ikelihood *e*stimation of familial *cor*relations has been implemented in a FORTRAN computer program, MLECOR, which is based on an efficient numerical optimization method (Lalouel 1979). The software is presently being generalized to incorporate nonrandom sampling. Copies of the program and documentation are available from the authors.

APPLICATIONS

Data on total serum cholesterol (CH) and HDL levels from the Swedish Twin Family Study (Crumpacker et al. 1979; Dahlén et al. 1983) are analyzed here to obtain maximum-likelihood estimates of familial correlations by using MLECOR. The study included adult monozygotic (MZ) and dizygotic (DZ) twin pairs born between 1911 and 1935 who were currently married and who had at least one adult child $\geq$18 years of age. Families of 39 male MZ, 37 female MZ, 32 male DZ, and 30 female DZ twin pairs were included in the sample. Only the 863 individuals for whom data were available on both CH and HDL were retained for analysis. The CH and HDL levels were determined using techniques described elsewhere (Dahlén et al. 1983). Each phenotype was adjusted for the effects of age, sex, and contraceptive usage as follows: Within each of five groups (male parents, female parents, male offspring, female offspring taking oral contraceptives, and female offspring not taking oral contraceptives), each phenotype was regressed on up to a third-degree polynomial in age, and those showing significant age effects were adjusted. The phenotypes were then standardized within each of the five groups. Finally, to minimize departures from normality, each phenotype was normalized using an inverse normal transformation of the ranks (Blom 1958).

Univariate Analysis of Nuclear Families

The data for families of twins can be used to illustrate several types of analyses. The simplest type of analysis requires splitting the families of twins into component nuclear families, such that one member of a twin pair, his or her spouse, and their offspring constitute one family and the co-twin, spouse, and offspring constitute a second family. As discussed earlier, in a given family the father alone represents the first group, the mother alone represents the second group, and all children constitute the third group. Some families may be devoid of data on one or more groups.

In this simple data structure, means for the father and mother (the elements of μ_1 and μ_2) are fixed at their sample values, as are the parental variances (the square of the diagonal elements of σ_1 and σ_2). There are six estimated parameters: the offspring mean (μ_3), offspring variance (square of σ_3), spouse correlation (element r_{12} in the matrix $\mathbf{R}$), the father-child and mother-child correlations

TABLE 1

HDL: ESTIMATES ± SE OF FAMILIAL CORRELATIONS, MEANS, AND VARIANCES, USING NUCLEAR-FAMILY DATA STRUCTURE

	Father	Mother	Child 1	Child 2
			A. **R**	
Father	1.00			
Mother	.20 ± .06	1.00		
Child 1	.30 ± .05	.30 ± .05	1.00	
Child 2	.30 ± .05[a]	.30 ± .05[a]	.29 ± .09	1.00
			B. μ	
	.01[b]	.00[b]	−.04 ± .06	−.04 ± .06[a]
			C. Variances (elements of σ squared)	
	0.79[b]	1.04[b]	1.12 ± 0.09	1.12 ± 0.09[a]

[a] Parameter constrained to equal the corresponding parameter for child 1.
[b] Fixed at sample value.

(r_{1j} and r_{2j}, respectively, in **R**, where j denotes an offspring), and sibling correlation (r_{jk}, where j and k denote two siblings).

For HDL, the maximum-likelihood estimates of the parameters are presented in table 1. Although data on all families with variable sibship sizes were analyzed together, the parameter estimates are shown for the special case of a two-child nuclear family, since this contains all the necessary information to describe families of any sibship size. The results indicate that the spouse correlation for HDL is .2; and the father-child, mother-child, and sibling correlations are remarkably similar, having maximum-likelihood estimates of ~.3.

Parameter estimates for CH are presented in table 2. There is no evidence for marital resemblance for CH levels, as indicated by the estimate of zero for the spouse correlation. The father-child (.22) and mother-child correlations (.21) are similar in magnitude, and the sibling correlation (.28) is somewhat higher.

Multivariate Analysis of Nuclear Families

Maximum-likelihood estimates from a bivariate analysis of HDL and CH levels in nuclear families are summarized in table 3. The estimates for each phenotype are essentially identical to those obtained previously for the univariate case. The between-phenotype estimates indicate that there are small phenotypic correlations in the parents but not in the offspring, and there is no evidence for cross-trait spouse, parent-offspring, or sibling resemblance.

Univariate Analysis of Families of Twins

The simultaneous analysis of the families of MZ and DZ twins provides more information for the resolution of genetic and environmental sources of familial

TABLE 2

CH: ESTIMATES ± SE OF FAMILIAL CORRELATIONS, MEANS, AND VARIANCES, USING NUCLEAR-FAMILY DATA STRUCTURE

	Father	Mother	Child 1	Child 2
			A. R	
Father	1.00			
Mother	.00 ± .06	1.00		
Child 1	.22 ± .06	.21 ± .05	1.00	
Child 2	.22 ± .06[a]	.21 ± .05[a]	.28 ± .08	1.00
			B. μ	
	.00[b]	.00[b]	−.01 ± .06	−.01 ± .06[a]
			C. Variances (elements of σ squared)	
	0.95[b]	1.02[b]	1.01 ± 0.08	1.01 ± 0.08[a]

[a] Parameter constrained to equal the corresponding parameter for child 1.
[b] Fixed at sample value.

resemblance and maternal effects (see, e.g., Nance et al. 1978; Williams and Iyer 1981; McGue et al. 1985) than does the analysis of nuclear families. This involves the simultaneous estimation of parameters in four types of families: those having (1) male MZ twins, (2) female MZ twins, (3) male DZ twins, or (4) female DZ twins. Since there are six groups in each type (twin 1, twin 2, spouse of twin 1, spouse of twin 2, children of twin 1, and children of twin 2), the maximum number of parameters is 132. However, of the 24 means and 24 variances, all but the pooled mean and variance for offspring (which require estimation) are again fixed at the sample values. Simultaneous estimation of the other means and variances has never given rise to visible differences either in parameter estimates or in the likelihood values (see, e.g., Rao et al. 1984). Of the 84 possible correlations, only 22 are nonredundant in a design in which the twins (but not the offspring) are broken down by sex: father-child, mother-child, sibling, MZ and DZ twin, MZ aunts and uncles, DZ aunts and uncles, spouse of MZ aunts and uncles, spouse of DZ aunts and uncles, cousins through MZ and DZ males and females, marital, MZ and DZ spouses of co-twins, and spouses of MZ and DZ twins. Consequently, only 24 parameters are estimated.

Results for the analysis of HDL, presented in table 4 (in tabular rather than matrix form, for simplicity), illustrate the multiple group analysis for a single variable. The pattern of correlations indicates the existence of strong familial resemblance for HDL. However, the data structure is too complex for ready interpretation based on the correlation estimates alone. In a separate report, McGue et al. (1985) analyzed these data by imposing a causal model, in which

TABLE 3

HDL AND CH SIMULTANEOUSLY: ESTIMATES ± SE OF FAMILIAL CORRELATIONS, MEANS, AND VARIANCES, USING NUCLEAR-FAMILY DATA STRUCTURE

	FATHER		MOTHER		CHILD 1		CHILD 2	
	HDL	CH	HDL	CH	HDL	CH	HDL	CH
				A. **R**				
Father:								
HDL	1.00							
CH	.14 ± .06	1.00						
Mother:								
HDL	.20 ± .06	.06 ± .06	1.00					
CH	−.03 ± .06	.00 ± .06	.15 ± .06	1.00				
Child 1:								
HDL	.29 ± .05	.03 ± .06	.31 ± .05	−.11 ± .06	1.00			
CH	−.05 ± .06	.22 ± .06	.12 ± .06	.20 ± .06	.02 ± .06	1.00		
Child 2:								
HDL	.29 ± .05[a]	.03 ± .06[a]	.31 ± .05[a]	−.11 ± .06[a]	.27 ± .09	.03 ± .07	1.00	
CH	−.05 ± .06[a]	.22 ± .06[a]	.12 ± .06[a]	.20 ± .06[a]	.03 ± .07[c]	.27 ± .09	.02 ± .06[a]	1.00
				B. μ				
	.00[b]	.01[b]	.00[b]	.00[b]	−.01 ± .05	−.04 ± .05	−.01 ± .05[a]	−.04 ± .05[a]
				C. Variances (elements of σ squared)				
	.95[b]	.79[b]	1.02[b]	1.04[b]	1.01 ± 0.08	1.11 ± 0.09	1.01 ± 0.08[a]	1.11 ± 0.09[a]

[a] Parameter constrained to equal the corresponding parameter for child 1.
[b] Fixed at sample value.
[c] Correlation of CH in child 1 with HDL in child 2 equals that of CH in child 2 with HDL in child 1.

TABLE 4

HDL: ESTIMATES ± SE OF FAMILIAL CORRELATIONS AND MEANS AND VARIANCES, USING FAMILIES-OF-TWINS DATA STRUCTURE

PARAMETER	Families of MZ Twins	Families of DZ Twins
Correlations:		
Marital[a]	.19 ± .08	.19 ± .08
Father-offspring[a]	.27 ± .08	.27 ± .08
Mother-offspring[a]	.31 ± .07	.31 ± .07
Siblings[a]	.29 ± .12	.29 ± .12
Twins	.72 ± .07	.38 ± .15
Cognate uncle–niece/nephew	.31 ± .10	.20 ± .15
Cognate aunt–niece/nephew	.38 ± .10	.35 ± .19
Affine uncle–niece/nephew	.09 ± .13	.10 ± .22
Affine aunt–niece/nephew	.00 ± .12	.03 ± .16
Cousins through male twins	.43 ± .13	−.06 ± .22
Cousins through female twins	.01 ± .15	.12 ± .28
Twin and co-twin's spouse	.28 ± .10	.14 ± .13
Spouses of twins	.05 ± .15	.20 ± .20
Means (Variances):		
Male twin[b]	−.12 (.74)	.02 (.78)
Female twin[b]	.24 (.83)	.24 (1.19)
Spouse of male twin[b]	−.23 (.89)	−.23 (1.07)
Spouse of female twin[b]	.02 (.64)	.16 (1.01)
Offspring[a]	−.01 ± .08 (1.13 ± 0.13)	−.01 ± .08 (1.13 ± 0.13)

[a] Parameters in MZ and DZ families are constrained to be equal.
[b] Parameters are fixed at sample values.

the correlations are expressed in terms of the model parameters, for the sources of familial resemblance.

The examples emphasize the general applicability of the method. The univariate analysis represents the case in which variable sample size exists within classes of relatives; the bivariate analysis of nuclear families illustrates the estimation of correlations when multiple phenotypes are measured within individuals in each class of relatives; and the example for families with twins illustrates a multiple group analysis, in which not all correlations within a group are independent of those in other groups.

DISCUSSION

The maximum-likelihood estimation procedure described here enables estimation of familial correlations when there is no closed-form solution, such as when pedigrees are not all of identical structure. Although application of maximum-likelihood estimation to such data structures has been described as presenting formidable problems (Karlin et al. 1981), the applications presented here and elsewhere (McGue et al. 1985; Vogler et al. 1987) illustrate that the numerical problems are not intractable. In fact, these methods were used in an extensive simulation experiment that involved maximum-likelihood estimation

of correlations from 18,000 separate samples of nuclear families, and in no case did they encounter numerical problems or fail to converge (Wette et al., submitted). Not only is maximum-likelihood estimation feasible, but the estimators of correlations, especially their z-transformations (Fisher 1921), were also shown to have desirable asymptotic properties, including normality.

Analysis of multivariate pedigree data necessarily involves the simultaneous estimation of many parameters, which gives rise to questions concerning the numerical problems associated with such estimation. Although these concerns cannot be taken lightly, it should be observed that our ability to successfully analyze complex data structures depends on several considerations: the information available in the data, the optimization method used, the computer facility available, and the type of model and parameterization used. Although the number of parameters may seem daunting in multivariate analyses, the information content also increases rapidly as the number of variables increases. For low-order multivariate problems, this counterbalances some of the potential problems in estimating large numbers of parameters. For example, although the 22 parameters of the bivariate application represent a substantial increase over the six parameters in each of the two univariate analyses, the SEs of relevant estimates are remarkably comparable in the two sets of analyses. If sample sizes are large enough to provide sufficient power for hypothesis testing, recent evidence indicates that multivariate analyses of the type discussed in the present paper are not problematical. For example, a trivariate analysis of lipoprotein fractions in 160 nuclear families provided sufficient power to test hypotheses and to estimate 31 parameters without encountering numerical difficulties (Vogler et al. 1987). Part of this success can be attributed to the optimization methods currently available for use. In the past, estimation of even 10 parameters simultaneously was problematical, but recent optimization packages such as GEMINI (Lalouel 1979), MAXLIK (Kaplan and Elston 1972), MINUIT (CERN 1977), and NAG (Numerical Algorithms Group 1978) have substantially improved our capability for simultaneous estimation of multiple parameters. Similarly, whereas in the past available computing facilities have limited the size of the problems that we could consider, recent advances in computing hardware are rapidly eliminating this limitation. Finally, the ability to estimate multiple parameters depends on the type of model and parameterization used. For example, segregation analysis of qualitative traits under a mixed model (Morton and MacLean 1974) depends on relatively subtle properties of the data used for resolution—and consequently might rapidly encounter numerical problems as more parameters are added. Analysis of the covariance structure of quantitative data under a multivariate normal model, as espoused in the present paper, relies on more robust properties of the data and is less likely to encounter numerical problems if, prior to attempting model fitting, the investigator is careful to scrutinize the data for violations of the distributional assumptions. Nevertheless, especially in multiparameter situations, one should make prudent use of the principle of parsimony and avoid modeling for its own sake. For example, fixing parental means and variances at sample values can consider-

ably reduce the number of parameters to be estimated simultaneously, thus improving numerical stability without sacrificing biologically meaningful information. In our experience, this has never led to appreciable changes either in parameter estimates or in likelihood values (although theoretically oriented statisticians might object to this practice in principle).

Overall, methodological, analytical, computational, and empirical considerations indicate that multiparameter multivariate analyses are feasible, within reasonable limits. Clearly, reliability of inference will depend on the size of the sample, the number of parameters, and the type of model used, and these issues will determine at what point numerical difficulties become a serious consideration. Although the analyses reported in the present paper presented no difficulties, there is clearly a need for systematic work to provide general guidelines on these issues.

This presentation represents a flexible framework for modeling generalized interclass and intraclass covariance structures in pedigrees. The proposed methodology can be readily extended or modified for other purposes. For example, although our presentation was confined to random sampling, the methods of Rao and Wette (in press) can be implemented for analysis of nonrandomly ascertained pedigree data. Whereas the phenotypic data were assumed to be adjusted for the effects of concomitant variables (e.g., age and sex) prior to estimation of the correlations, the means and variances may instead be modeled as functions of the concomitant variables if desired. If the intraclass and interclass assumptions seem to be inappropriate for particular situations, they may be selectively relaxed. However, the methods can be used as proposed to achieve certain specific objectives. For example, if sex-specific sibling or adult-versus-pediatric-sibling correlations are desired, they can be accomplished by suitably defining the groups. Finally, although we did not explicitly discuss tests of significance, likelihood-ratio tests of hypotheses can be carried out using the proposed methodology. One may test for the significance of a particular correlation or for the equality of several correlation coefficients.

For a particular application, the complexity of the analysis depends on the information that one wants to extract from the data. Univariate nuclear-family analyses quantify the magnitude of familial resemblance for one phenotype. Multivariate nuclear-family analyses, as well as providing the same information as univariate analyses of each phenotype separately, also illustrate the nature of covariation among phenotypes both within and among individuals. More complex analyses, such as those conducted using the family-of-twins design, enable the estimation of correlations that arise from more unusual familial relationships—such as twinship and the various aunt/uncle–niece/nephew and cousin relationships—in addition to the basic nuclear-family correlations. Such complex analyses are useful for testing alternate models for the sources of familial resemblance. Thus, depending on the information to be extracted from the data, the maximum-likelihood method can be applied to a data structure of any complexity, within the limits of the number of parameters that can be estimated using an iterative procedure.

ACKNOWLEDGMENTS

This work was partly supported by grants GM28719, HL33973, MH17104, and MH31302 from the National Institutes of Health and National Institute of Mental Health.

REFERENCES

Blom, G. 1958. Statistical estimates and transformed beta variables. Wiley, New York.

CERN. 1977. MINUIT: a system for function minimization and analysis of the parameter errors and correlations. CERN, Geneva.

Crumpacker, D. W., R. Cederlöf, L. Friberg, W. J. Kimberling, S. Sörensen, S. G. Vandenberg, J. S. Williams, G. E. McClearn, B. Grever, H. Iyer, M. J. Krier, N. L. Pedersen, R. A. Price, and I. Roulette. 1979. A twin methodology for the study of genetic and environmental control of variation in human smoking behavior. Acta Genet. Med. Gemellol. (Roma) **28**:173–195.

Dahlén, G., C. Ericson, U. deFaire, L. Iselius, and T. Lundman. 1983. Genetic and environmental determinants of cholesterol and HDL-cholesterol concentrations in blood. Int. J. Epidemiol. **12**:32–35.

Donner, A., and J. J. Koval. 1980. The estimation of intraclass correlation in the analysis of family data. Biometrics **36**:19–25.

Fisher, R. A. 1921. On the probable error of a coefficient of correlation deduced from a small sample. Metron **1**:3–32.

Kaplan, E. B., and R. C. Elston. 1972. A subroutine package for maximum likelihood estimation (MAXLIK). Inst. Stat. Mimeo Ser. 823. University of North Carolina, Chapel Hill.

Karlin, S., E. C. Cameron, and P. T. Williams. 1981. Sibling and parent-offspring correlation estimation with variable family size. Proc. Natl. Acad. Sci. USA **78**:2664–2668.

Lalouel, J. M. 1979. GEMINI—a computer program for optimization of general nonlinear functions. Univ. Utah Tech. Rep. 14.

McGue, M., D. C. Rao, L. Iselius, and J. M. Russell. 1985. Resolution of genetic and cultural inheritance in twin families by path analysis: application to HDL-cholesterol. Am. J. Hum. Genet. **37**:998–1014.

Morton, N. E., and C. J. MacLean. 1974. Analysis of family resemblance. III. Complex segregation analysis of quantitative traits. Am. J. Hum. Genet. **26**:489–503.

Nance, W. E., L. A. Corey, and J. A. Boughman. 1978. Monozygotic twin kinships: a new design for genetic and epidemiologic research. Pp. 87–132 *in* N. E. Morton and C. S. Chung, eds. Genetic epidemiology. Academic Press, New York.

Numerical Algorithms Group. 1978. EO4 JAF. Mark 8. Numerical Algorithms Group, Oxford.

Rao, D. C., P. M. Laskarzewski, J. A. Morrison, P. Khoury, K. Kelly, R. Wette, J. M. Russell, and C. J. Glueck. 1982. The Cincinnati Lipid Research Clinic Family Study: cultural and biological determinants of lipids and lipoprotein concentrations. Am. J. Hum. Genet. **34**:888–903.

Rao, D. C., M. McGue, R. Wette, and C. J. Glueck. 1984. Path analysis in genetic epidemiology. Pp. 35–81 *in* A. Chakravarti, ed. Human population genetics: the Pittsburgh Symposium. Van Nostrand Reinhold, New York.

Rao, D. C., N. E. Morton, C. L. Gulbrandsen, G. G. Rhoads, A. Kagan, and S. Yee. 1979. Cultural and biological determinants of lipoprotein concentrations. Ann. Hum. Genet. **42**:467–477.

Rao, D. C., N. E. Morton, J. M. Lalouel, and R. Lew. 1979. Analysis of family resemblance. I. Path analysis of quantitative data under mixed homogamy. Univ. Hawaii PGL Tech. Rep. 2.

Rao, D. C., and R. Wette. Non-random sampling in genetic epidemiology: maximum likelihood methods for truncate selection. Genet. Epidemiol. (in press).

Rosner, B. 1979. Maximum likelihood estimates of interclass correlations. Biometrika **66**:533–538.

———. 1982. On the estimation and testing of interclass correlations: the general case of multiple replicates for each variable. Am. J. Epidemiol. **116**:722–730.

Vogler, G. P., D. C. Rao, P. M. Laskarzewski, C. J. Glueck, and J. M. Russell. 1987. Multivariate analysis of lipoprotein cholesterol fractions. Am. J. Epidemiol. **125**:706–719.

Wette, R., M. McGue, D. C. Rao, and C. R. Cloninger. On the properties of maximum likelihood estimates of familial correlations under variable sibship size. Biometrics (submitted).

Williams, J. S., and H. A. Iyer. 1981. A statistical model and analysis for genetic and environmental effects in responses from twin family studies. Acta Genet. Med. Gemellol. (Roma) **30**:9–38.